Please cite this paper as:

Neuwirth, E. (2006), "PISA 2000: Sample Weight Problems in Austria", *OECD Education Working Papers*, No. 5, OECD Publishing.
doi:10.1787/220456725273

OECD Education Working Papers
No. 5

PISA 2000

SAMPLE WEIGHT PROBLEMS IN AUSTRIA

Erich Neuwirth

Unclassified **EDU/WKP(2006)1**

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development **09-Jun-2006**

English - Or. English

DIRECTORATE FOR EDUCATION

EDU/WKP(2006)1
Unclassified

PISA 2000: Sample Weight Problems in Austria

(Education Working Paper No. 5)

About the author

Erich Neuwirth is a member of the Working Group for Data Analysis and Computing at the Faculty of Computer Science of the University of Vienna. He teaches computer science, statistics and mathematics, and is the head of the Center for Didactics of Computer Science. One of his research areas is the Rasch model which is the psychometric foundation of the PISA methodology.

Mr. Erich Neuwirth
Faculty of Computer Science, University of Vienna
Universitaetsstr. 5-9, Vienna, Austria, A-1010

JT03210368

English - Or. English

OECD Education Working Papers

This series is designed to make available to a wider readership selected studies on education prepared for use within the OECD. Authorship is usually collective, but principal writers are named. The papers are generally available only in their original language English or French with a summary in the other.

Comment on the series is welcome, and should be sent to either edu.contact@oecd.org or the Directorate for Education, 2, rue André Pascal, 75775 PARIS CEDEX 16, France.

The opinions expressed in these papers are the sole responsibility of the author(s) and do not necessarily reflect those of the OECD or of the governments of its member countries.

---

www.oecd.org/edu/workingpapers

INTRODUCTORY NOTE

As noted in the OECD's *PISA 2000 Technical Report* (OECD, 2002), the Austrian sample for the PISA 2000 assessment did not adequately cover students enrolled in combined school and work-based vocational programmes as required by the OECD's technical standards for PISA.

The purpose of this working paper is to quantify the comparability problems resulting from the inadequate coverage of the PISA target population in the Austrian PISA 2000 assessment and to establish adjustments that could be used to correct for this and thus to allow reliable comparisons between the 2000 and 2003 data. Using the supplementary data for the number of students in the PISA strata provided by the Austrian Ministry Education, this report presents adjusted student weights for analysing the PISA 2000 Austrian data.

Comparing Austria's scores in reading, mathematics and science for PISA 2000 and PISA 2003 reveals noticeable differences in performance:

	Reading	Mathematics	Science
PISA 2000	507.1	515.0	518.6
PISA 2003	490.7	505.6	491.0
Difference	-16.4	-9.4	-27.7

Table 1: Austria's country mean scores for three domains in PISA 2000 and PISA 2003

An analysis by gender shows very surprising results:

	Reading		Mathematics		Science	
	Female	Male	Female	Male	Female	Male
PISA 2000	520.3	494.6	503.0	530.1	513.9	525.7
PISA 2003	514.4	467.1	501.8	509.4	492.3	489.7
Difference	-5.9	-27.5	-1.2	-20.7	-21.6	-36.0

Table 2: Austria's mean scores for three domains by gender in PISA 2000 and PISA 2003

According to Table 2, the changes for male students are much larger than for female students. Since the PISA sample is a stratified sample, it is worthwhile to look at the stratification methods used in PISA and to investigate whether dissimilarities in strata percentages are responsible for at least part of these differences.

WEIGHTS AND STRATA

In Austria, the PISA 2000 target population was classified into 19 strata according to the type of school the students attended. Country scores are computed as weighted means of individual values, and the weights should, in theory, reflect the proportions of the strata within the target population. The following table compares total weights and percentages for the strata in the Austrian student population for the PISA samples 2000 and 2003.

Long name	Short name	Sample weight		Weight percentage	
		PISA 2000	PISA 2003	PISA 2000	PISA 2003
Hauptschule	HS	2518.1	3756.4	3.6%	4.4%
Polytechnische Schule	Poly	6829.7	7447.1	9.7%	8.7%
Sonderschule	SoS	277.5	783.7	0.4%	0.9%
Gymnasium	G	6659.2	7324.9	9.4%	8.5%
Realgym und WkGym	RG+WkG	5871.5	5669.4	8.3%	6.6%
ORG	ORG	4329.8	5297.2	6.1%	6.2%
AHS sonst	AHSdiv	285.0	0.0	0.4%	0.0%
BS tech/gew	BStg	5460.8	11561.7	7.7%	13.5%
BS kaufm	BSk	2854.7	5418.1	4.0%	6.3%
BS land/fw	BSlf	85.4	0.0	0.1%	0.0%
BMS gew/tech	BMSgt	2625.0	2059.9	3.7%	2.4%
BMS kaufm	BMSk	2513.9	3880.4	3.6%	4.5%
BMS wiso	BMSws	2826.0	3738.3	4.0%	4.4%
BMS lw/fo	BSMlf	2727.2	3775.5	3.9%	4.4%
BHS tech	BGSt	10572.7	10476.5	14.9%	12.2%
BHS kauf	BHSk	7755.1	7480.8	11.0%	8.7%
BHS wi/so	BHSws	4395.5	5587.7	6.2%	6.5%
BHS lw/fw	BHSlf	658.5	386.7	0.9%	0.5%
Lehrer/ErzieherBld	LeErz	1482.0	1286.7	2.1%	1.5%
Total		70727.6	85931.0	100.0%	100.0%

Table 3: Sample weights for PISA 2000 and PISA 2003 by stratum

In PISA, the sum of weights for each country should give the size of the target population. For the Austrian data, the sum of weights increased from PISA 2000 to 2003 by approximately 20% (85931.0/70727.6=1.21). It is totally implausible that the target population in Austria should have changed by this amount within three years.

Additionally, strata percentages for PISA 2000 and 2003 are highly different. The most noticeable difference is for BS tech/gew (Technisch-gewerbliche Berufsschule, BStg) stratum, which was 7.7% in PISA 2000 and 13.5% in PISA 2003. This stratum has a very high percentage of male students. Since in general the gender ratio in Austria is quite different for different school types, it seems advisable to inspect the table of sample weights classified by gender and stratum. Table A1 in the appendix shows these numbers.

Table 3 above also shows that two strata (AHSdiv and BSlf) had very low weight in PISA 2000; therefore they were joined with their respective most similar strata in PISA 2003.

In addition to the weights from the PISA 2003 sample, the author also received access to Austrian school statistics from the Austrian Ministry of Education. Table A2 gives student numbers by stratum and gender for the school years 1999-2000 and 2002-2003.

The idea of stratified sampling is that the total weights for each group defined by stratum and gender should be similar to the corresponding numbers from the population. Tables A2 show that the distribution of the population is quite similar for 2000 and 2003, but Table A1 shows that the distribution in the samples differs noticeably. Figures 1 and 2 illustrate this fact for female and male students, respectively.

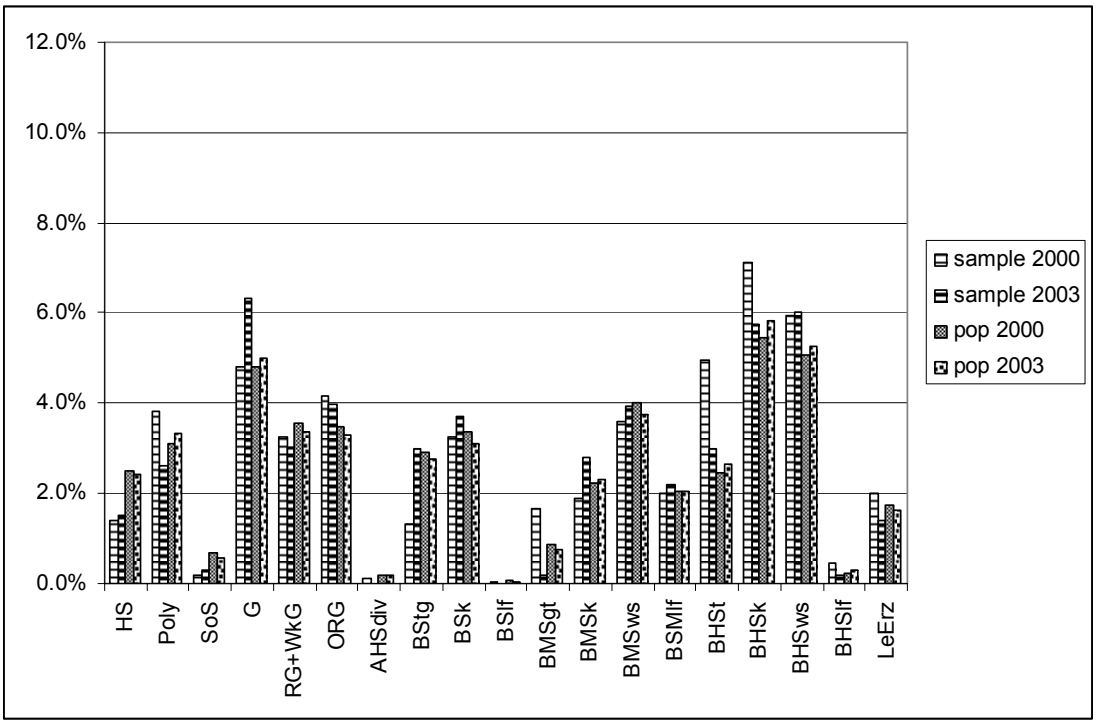


Figure 1: Stratum weight percentages, female students

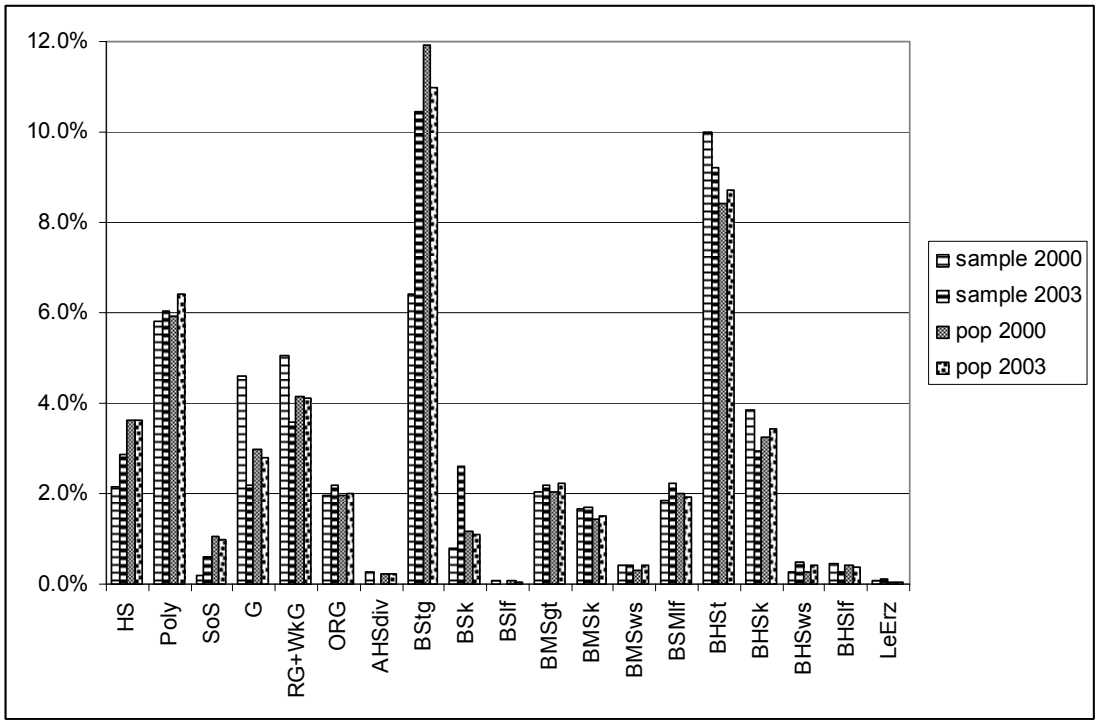


Figure 2: Stratum weight percentages, male students

Figure 1 and 2 show that male students in the BStg stratum were clearly underrepresented in the PISA 2000 sample. On the other hand, students in the BHSt stratum were slightly overrepresented in PISA 2000 sample. It is worth noticing that students in the BStg stratum, where students are apprentices attending vocational schools, tended to perform low and students in the BHSt stratum tended to perform high (see Table 4). Regarding female students, students in the BHSk stratum were overrepresented in the PISA 2000 sample and students in the G stratum were overrepresented in the PISA 2003 sample. Students in the both strata tended to perform high.

WEIGHT INFLUENCE ON SCORES

Academic performance varies widely with school type and gender. Therefore, misrepresenting some strata in the sample may severely bias the overall score for the country. Table 4 gives the (weighted) means for reading, mathematics and science by stratum and gender for the PISA 2000 sample. This table shows quite large differences for stratum and gender, especially for reading. Note that the male students in the BStg stratum have a low reading score. Therefore, the fact that they are underrepresented in the PISA 2000 sample leads to a too high country average score for Austria in PISA 2000.

Stratum	Reading		Mathematics		Science	
	Female	Male	Female	Male	Female	Male
HS	356.3	358.3	365.7	396.7	357.5	395.3
Poly	427.1	412.6	417.4	450.0	425.8	458.6
SoS	291.1	299.7	291.9	331.6	325.0	352.1
G	591.6	557.6	563.7	578.9	584.0	590.3
RG+WkG	564.9	555.2	558.3	590.1	553.0	582.3
ORG	559.8	528.3	528.1	554.2	553.9	554.8
AHSdiv	463.1	479.0	458.5	506.4	452.0	484.1
BStg	447.2	437.1	428.0	481.5	438.7	464.2
BSk	456.2	462.6	434.9	466.2	447.0	470.0
BSlf	428.6	365.5	462.9	434.6	418.5	420.8
BMSgt	472.7	490.4	458.5	532.6	465.6	524.7
BMSk	496.1	483.8	480.3	510.5	479.1	510.6
BMSws	489.1	488.7	469.5	498.7	477.2	525.1
BSMlf	473.0	451.4	466.9	487.5	469.8	473.5
BGSt	558.7	535.3	541.0	580.1	554.3	566.5
BHSk	549.7	550.6	532.0	568.7	541.8	583.5
BHSws	540.7	546.9	519.2	538.5	535.1	547.2
BHSlf	569.6	546.4	579.1	609.3	572.7	578.4
LeErz	562.0	560.6	546.5	606.3	547.0	568.9

Table 4: Domain scores by gender and stratum for PISA 2000

To study the influence of weighting on the overall country score, two new weighting schemes should be created by multiplying all weights within one cell defined by stratum and gender with the same constant. These newly created weights can be applied to the PISA 2000 data to calculate the Austrian country mean scores for three domains in order to examine how Austria's overall scores vary according to different weighting schemes For the first newly created weights, namely "sample2003", the constants are chosen such that the weighted cell sums would be equal to the corresponding weighted cell sums of the PISA 2003 sample. The second weights, namely "pop2000", give each cell a total weight equal to the number of students for that cell in the target population. The two strata not explicitly used in PISA 2003

are received zero total weight for the "sample2003" weights. Since the total weight of both strata in PISA 2000 was only 0.5%, this effect seems negligible.

Table 5 shows, based on the PISA 2000 data, Austria's country mean scores for reading, mathematics and science using three different kinds of weighting schemes – the original PISA 2000 weights, namely "orig2000" as well as two newly created weights of "sample2003" and "pop2000". Table 2.3 shows the scores by gender with three different weighting schemes based on PISA 2000 data.

	With the weights of:	Reading	Mathematics	Science
PISA 2000	"orig2000"	507.1	515.0	518.6
	"sample2003"	498.4	506.9	509.9
	"pop2000"	492.1	502.5	504.7

Table 5: Austria's mean scores for three domains with three different weighting schemes using the PISA 2000 data

	With the weights of:	Reading		Mathematics		Science	
		Female	Male	Female	Male	Female	Male
PISA 2000	"orig2000"	520.3	494.6	503.0	530.1	513.9	525.7
	"sample2003"	518.3	478.6	500.0	513.8	510.8	509.0
	"pop 2000"	509.2	475.8	492.5	512.0	502.2	507.1

Table 6: Austria's mean scores by gender for three domains with three different weighting schemes using the PISA 2000 data

From Table 6, it can be seen that both "sample2003" and "pop2000" weights schemes produce very similar scores for male students. For the female students, however, mean scores with "sample2003" tend to be higher than those with "pop2000" weights. This could be because female students in the G stratum are overrepresented in the PISA 2003 sample and because these students tend to be high achievers.

In an internal preliminary report prepared by Neuwirth *et al.*[1] for submission to the Austrian Ministry of Education, the PISA 2000 data were recomputed with the "sample2003" weights in order to show that part of the decline in Austria's performance in PISA reported by the media was simply due to the biased student sampling procedure used in Austria for PISA 2000. The OECD's PISA publications are aware of the problem. In the OECD's *PISA 2000 Technical Report* (OECD, 2002), the country comment about Austria states:

> Thus, it is not possible to assess how well the students sampled from vocational schools represent the universe of students enrolled in vocational schools, and so those students not attending classes at the time of the PISA assessment are not represented in the PISA result*s*.

[1] This internal report was written by Erich Neuwirth and Wilfried Grossmann (Institute for Scientific Computing, University of Vienna), Ivo Ponocny (Statistik Austria), and Peter Steiner (IHS Wien).

At the time of the preliminary report, the official data about the school statistics were not available broken down by gender, therefore the report could not use the "pop2000" weights. However, it is clear that the best way of adjusting the misrepresentation of the strata in the PISA 2000 sample is to use the "pop2000" weights created based on the Austrian school statistics of the Austrian Ministry of Education. Therefore, it is recommended to use the "pop2000" weights in order to compare the Austrian PISA 2000 results with other countries and also with the Austrian PISA 2003 results. For example, if the results in Table 7 and Table 8 are compared with those in Table 1 and Table 2, the differences in scores between PISA 2000 and PISA 2003 decrease when the corrected weights are applied for the PISA 2000 data, even though significant decline remains in the science domain. With the corrected weights for PISA 2000, all analyses can be performed using standard PISA methodology by adjusting the 80 Balanced Repeated Replication (BRR) variables according to the corrected weights.

	Reading	Mathematics	Science
PISA 2000 with "pop2000" weights	492.1	502.5	504.7
PISA 2003	490.7	505.6	491.0
Difference	-1.4	3.1	-13.7

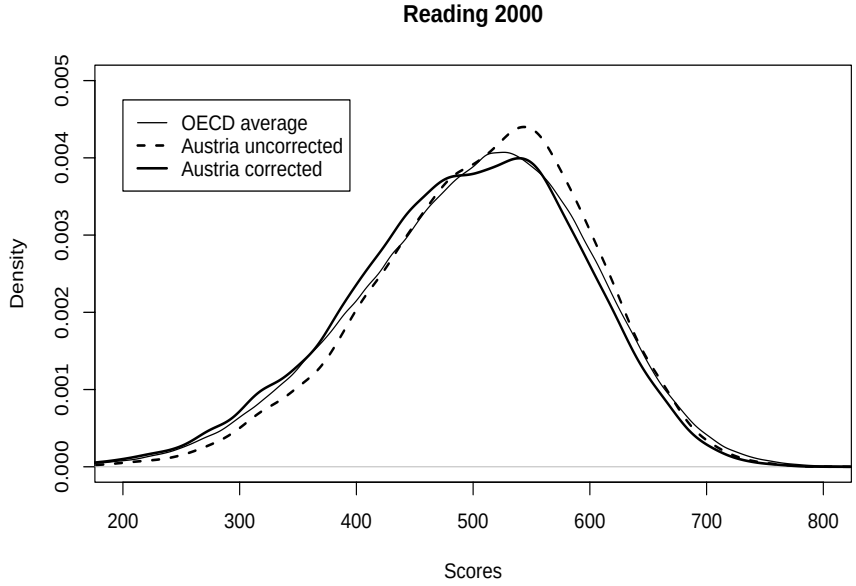
Table 7: Austria's country mean scores for three domains for PISA 2000 with the corrected weights and PISA 2003

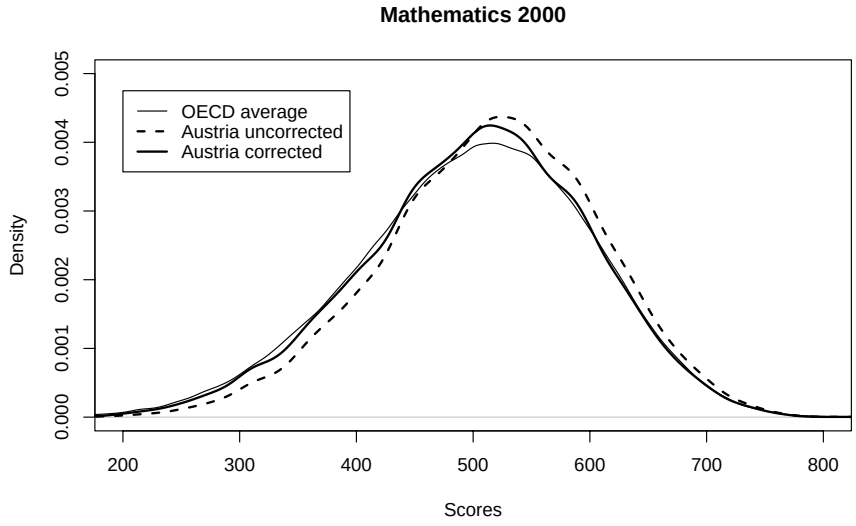
	Reading		Mathematics		Science	
	Female	Male	Female	Male	Female	Male
PISA 2000 with "pop2000" weights	509.2	475.8	492.5	512.0	502.2	507.1
PISA 2003	514.4	467.1	501.8	509.4	492.3	489.7
Difference	5.2	-8.7	9.3	-2.6	-9.9	-17.4

Table 8: Austria's mean scores by gender for three domains for PISA 2000 with the corrected weights and PISA 2003

Statistical analyses should not stop at computing means. Therefore, density estimators for the distribution of the plausible values for the student scores (with "pop2000"weights and – to visualize the corrections – with "orig2000"weights) should be computed. Since there are five plausible score values for each student for each domain, our density estimators use variables concatenating the five plausible value vectors. In this way, each student from the sample is represented by five values in the density plot. This is a simple implementation of the model assumption of PISA that each student is best represented by an *a posteriori* distribution instead of a single score.

Figure 3 below shows the difference of the distributions for the original weights ("orig2000") and the corrected weights ("pop2000"). For all three domains, the difference in the distributions is clearly visible. The OECD average distribution is the distribution of a data set including all 27 OECD countries participating fully in PISA 2000. Each of these 27 countries has the same total weight; therefore, all countries contribute equally to this distribution.


Reading 2000
OECD average
Austria uncorrected
Austria corrected
Density
0.000
0.001
0.002
0.003
0.004
0.005
200
300
400
500
600
700
800
Scores


Mathematics 2000
OECD average
Austria uncorrected
Austria corrected
Density
0.000
0.001
0.002
0.003
0.004
0.005
200
300
400
500
600
700
800
Scores

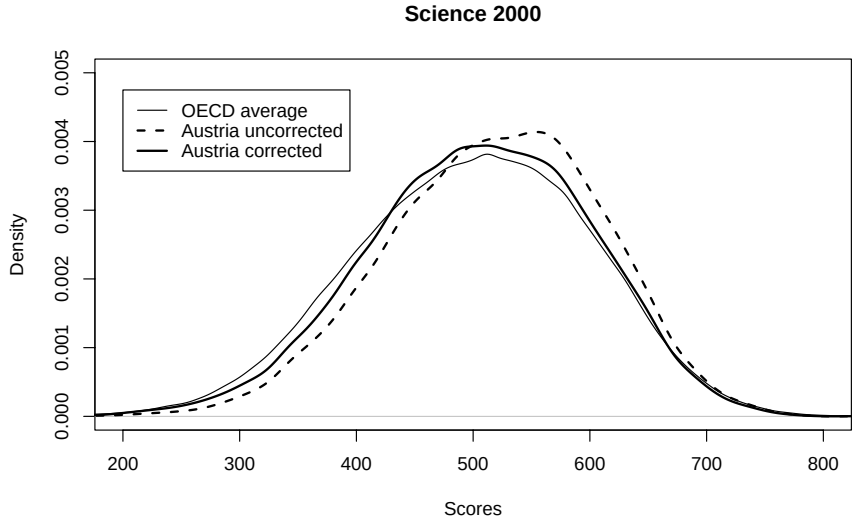


Figure 3: Distribution of plausible values for student scores for the PISA 2000 data with corrected weights ("pop2000") and uncorrected weights ("orig2000")

The Austrian score distribution using the corrected weights ("pop2000") is similar to the OECD average distribution. On the other hand, using the original weights ("orig2000") produces score distributions for Austria that seem to be better than average.

CONCLUSIONS

The Austrian results for PISA 2000 are biased because of the misrepresentation of a stratum of the target population in PISA 2000. The net effect of this misrepresentation is that the national average scores for PISA 2000 are higher than those calculated using correct stratum weights. Austrian results for PISA 2003 give a much more faithful impression of student achievements. Since the biased Austrian results for 2000 are too high, the Austrian PISA 2003 national report erroneously reported a decline in performance in all three PISA domains. As a consequence, modified sets of weights will be made available. This will allow for less biased longitudinal analyses, including Austrian data in international comparisons in future studies based on PISA data.

REFERENCES

Haider, Günter, and Claudia Reiter (Hrsg.) (2004), *PISA 2003: Internationaler Vergleich von Schülerleistungen,* Nationaler Bericht, Leykam, Graz.

OECD (2001), *Knowledge and Skills for Life: First Results from PISA 2000*, OECD, Paris.

OECD (2002), *PISA 2000 Technical Report,* OECD, Paris.

OECD (2004), Learning for Tomorrow's World – First Results from PISA 2003, OECD, Paris.

APPENDIX A

Table A1: Sample weights from PISA 2000 and PISA 2003, by stratum and gender

Stratum	Sample PISA 2000		Sample PISA 2003		Percentage by gender				Percentage total			
					PISA 2000		PISA 2003		PISA 2000		PISA 2003	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
HS	994.1	1524.0	1291.0	2465.3	2.7%	4.5%	3.0%	5.7%	1.4%	2.2%	1.5%	2.9%
Poly	2715.1	4114.5	2248.3	5198.8	7.4%	12.1%	5.2%	12.1%	3.8%	5.8%	2.6%	6.0%
SoS	138.7	138.7	266.9	516.8	0.4%	0.4%	0.6%	1.2%	0.2%	0.2%	0.3%	0.6%
G	3396.4	3262.7	5446.1	1878.8	9.2%	9.6%	12.7%	4.4%	4.8%	4.6%	6.3%	2.2%
RG+WkG	2300.8	3570.7	2597.7	3071.7	6.2%	10.5%	6.1%	7.1%	3.3%	5.0%	3.0%	3.6%
ORG	2951.0	1378.8	3417.2	1880.0	8.0%	4.1%	8.0%	4.4%	4.2%	1.9%	4.0%	2.2%
AHSdiv	91.3	193.7	0.0	0.0	0.2%	0.6%	0.0%	0.0%	0.1%	0.3%	0.0%	0.0%
BStg	924.2	4536.6	2569.2	8992.5	2.5%	13.4%	6.0%	20.9%	1.3%	6.4%	3.0%	10.5%
BSk	2306.6	548.1	3171.6	2246.5	6.3%	1.6%	7.4%	5.2%	3.3%	0.8%	3.7%	2.6%
BSlf	32.3	53.2	0.0	0.0	0.1%	0.2%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%
BMSgt	1188.6	1436.4	170.9	1888.9	3.2%	4.2%	0.4%	4.4%	1.7%	2.0%	0.2%	2.2%
BMSk	1329.7	1184.2	2420.3	1460.1	3.6%	3.5%	5.6%	3.4%	1.9%	1.7%	2.8%	1.7%
BMSws	2540.8	285.2	3369.3	369.0	6.9%	0.8%	7.9%	0.9%	3.6%	0.4%	3.9%	0.4%
BSMlf	1426.4	1300.8	1871.6	1903.9	3.9%	3.8%	4.4%	4.4%	2.0%	1.8%	2.2%	2.2%
BGSt	3505.7	7067.0	2560.1	7916.4	9.5%	20.9%	6.0%	18.4%	5.0%	10.0%	3.0%	9.2%
BHSk	5030.9	2724.2	4949.2	2531.6	13.7%	8.0%	11.5%	5.9%	7.1%	3.9%	5.8%	2.9%
BHSws	4212.7	182.8	5169.7	418.0	11.4%	0.5%	12.1%	1.0%	6.0%	0.3%	6.0%	0.5%
BHSlf	331.5	327.0	164.9	221.9	0.9%	1.0%	0.4%	0.5%	0.5%	0.5%	0.2%	0.3%
LeErz	1427.4	54.6	1191.6	95.1	3.9%	0.2%	2.8%	0.2%	2.0%	0.1%	1.4%	0.1%
Total	36844.3	33883.3	42875.8	43055.3	100.0%	100.0%	100.0%	100.0%	52.1%	47.9%	49.9%	50.1%

Table A2: Student numbers from the Austrian school statistics of the Austrian Ministry of Education, by stratum and gender

Stratum	Population in PISA 2000		Population in PISA 2003		Percentage by gender				Percentage total			
					PISA 2000		PISA 2003		PISA 2000		PISA 2003	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
HS	2239	3246	2172	3230	5.1%	7.0%	5.0%	7.0%	2.5%	3.6%	2.4%	3.6%
Poly	2804	5338	2990	5741	6.4%	11.6%	6.9%	12.5%	3.1%	5.9%	3.3%	6.4%
SoS	628	950	510	878	1.4%	2.1%	1.2%	1.9%	0.7%	1.1%	0.6%	1.0%
G	4316	2671	4488	2499	9.8%	5.8%	10.3%	5.4%	4.8%	3.0%	5.0%	2.8%
RG+WkG	3201	3737	3025	3691	7.3%	8.1%	6.9%	8.0%	3.6%	4.2%	3.4%	4.1%
ORG	3135	1766	2959	1804	7.1%	3.8%	6.8%	3.9%	3.5%	2.0%	3.3%	2.0%
AHSdiv	174	213	179	201	0.4%	0.5%	0.4%	0.4%	0.2%	0.2%	0.2%	0.2%
BStg	2636	10741	2465	9819	6.0%	23.3%	5.7%	21.4%	2.9%	11.9%	2.8%	11.0%
BSk	3036	1053	2780	974	6.9%	2.3%	6.4%	2.1%	3.4%	1.2%	3.1%	1.1%
BSlf	67	57	44	35	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%
BMSgt	769	1838	684	2000	1.8%	4.0%	1.6%	4.4%	0.9%	2.0%	0.8%	2.2%
BMSk	2005	1294	2077	1360	4.6%	2.8%	4.8%	3.0%	2.2%	1.4%	2.3%	1.5%
BMSws	3615	258	3361	370	8.2%	0.6%	7.7%	0.8%	4.0%	0.3%	3.8%	0.4%
BSMlf	1823	1797	1822	1724	4.2%	3.9%	4.2%	3.8%	2.0%	2.0%	2.0%	1.9%
BGSt	2205	7555	2360	7813	5.0%	16.4%	5.4%	17.0%	2.5%	8.4%	2.6%	8.7%
BHSk	4887	2910	5235	3073	11.1%	6.3%	12.0%	6.7%	5.4%	3.2%	5.8%	3.4%
BHSws	4549	250	4704	364	10.4%	0.5%	10.8%	0.8%	5.1%	0.3%	5.3%	0.4%
BHSlf	208	377	262	341	0.5%	0.8%	0.6%	0.7%	0.2%	0.4%	0.3%	0.4%
LeErz	1571	44	1441	43	3.6%	0.1%	3.3%	0.1%	1.7%	0.0%	1.6%	0.0%
Total	43868	46095	43558	45960	100.0%	100.0%	100.0%	100.0%	48.8%	51.2%	48.7%	51.3%

OECD EDUCATION WORKING PAPERS

No. 1 TEACHER DEMAND AND SUPPLY: IMPROVING TEACHING QUALITY AND ADDRESSING TEACHER SHORTAGES (2002), Paulo Santiago

No. 2 TEACHER EDUCATION AND THE TEACHING CAREER IN AN ERA OF LIFELONG LEARNING (2002), John Coolahan

No. 3 TOWARDS AN UNDERSTANDING OF THE MECHANISMS THAT LINK QUALIFICATIONS AND LIFELONG LEARNING, Friederike Behringer and Mike Coles

No. 4 MEASURING EDUCATIONAL PRODUCTIVITY IN STANDARDS-BASED ACCOUNTABILITY SYSTEMS: INTRODUCING THE SES RETURN ON SPENDING INDEX, Martin Hempel

OECD Education Working Papers can be found on the OECD Education Web site at: www.oecd.org/edu/workingpapers. If you wish to be informed about the release of new OECD Education working papers, please: (i) Go to the OECD Web site: www.oecd.org; (ii) Click on "My OECD"; (iii) Sign up and create an account with "My OECD"; (iv) Select "Education" as one of your favourite themes; and (v) Pick "OECD Education Working Papers" as one of the newsletters you would like to receive. For further information please write to: edu.contact@oecd.org.